

Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past

Clio Der Sarkissian, Vianney Pichereau, Catherine Dupont, Peter C Ilsøe, Mickael Perrigault, Paul Butler, Laurent Chauvaud, Jon Eiriksson, James Scourse, Christine Paillard, et al.

▶ **To cite this version:**

Clio Der Sarkissian, Vianney Pichereau, Catherine Dupont, Peter C Ilsøe, Mickael Perrigault, et al.. Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. Molecular Ecology Resources, 2017, 17 (5), pp.835-853. 10.1111/1755-0998.12679 . hal-01624435

HAL Id: hal-01624435

https://univ-rennes.hal.science/hal-01624435

Submitted on 26 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Article

PROF. LUDOVIC ORLANDO (Orcid ID : 0000-0003-3936-1850)

Article type : Resource Article

Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past

Clio Der Sarkissian[1], Vianney Pichereau[2], Catherine Dupont[3], Peter C. Ilsøe[1], Mickael Perrigault[2], Paul Butler[4], Laurent Chauvaud[2], Jón Eiríksson[5], James Scourse[4], Christine Paillard[2], Ludovic Orlando[1,6]

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen K 1350, Denmark;

[2]Lemar UMR 6539 CNRS/UBO/IRD/Ifremer, Université de Brest, IUEM, Rue Dumont d'Urville, 29280 Plouzané, France;

[3]CReAAH, UMR6566, Université de Rennes 1, Bâtiment 24-25 Campus de Beaulieu, 263 Avenue du Général Leclerc Campus de Beaulieu CS 74205, 35042 Rennes, France;

[4]CGES, College of Life and Environmental Sciences, University of Exeter, Penryn, Cornwall, TR10 9EZ, United Kingdom;

[5]Institute of Earth Sciences, University of Iceland, Askja, IS-101 Reykjavík, Iceland;

Accepted Article

[6]Université de Toulouse, University Paul Sabatier (UPS), Laboratoire AMIS, CNRS UMR 5288, 37 allées Jules Guesde, 31000 Toulouse, France.

KEYWORDS marine mollusc shells; ancient DNA; metagenomics; high-throughput DNA sequencing

CORRESPONDING AUTHOR

Ludovic Orlando; Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, Copenhagen K 1350, Denmark; Fax: +4535322325; E-mail: Lorlando@snm.ku.dk

RUNNING TITLE

Metagenomics of ancient marine mollusc shells

ABSTRACT

Marine mollusc shells enclose a wealth of information on coastal organisms and their environment. Their life history traits as well as (palaeo-) environmental conditions, including temperature, food availability, salinity, and pollution, can be traced through the analysis of their shell (micro-) structure and biogeochemical composition. Adding to this list the DNA entrapped in shell carbonate biominerals potentially offers a novel and complementary proxy both for reconstructing palaeoenvironments and tracking mollusc evolutionary trajectories. Here, we assess this potential by applying DNA extraction, high-throughput shotgun DNA sequencing and metagenomic analyses to marine mollusc shells spanning the last ~7,000 years. We report

Accepted Article

successful DNA extraction from shells, including a variety of ancient specimens, and find that DNA recovery is highly dependent on their biomineral structure, carbonate layer preservation, and disease state. We demonstrate positive taxonomic identification of mollusc species using a combination of mitochondrial DNA genomes, barcodes, genome-scale data, and metagenomic approaches. We also find shell biominerals to contain a diversity of microbial DNA from the marine environment. Finally, we reconstruct genomic sequences of organisms closely related to the *Vibrio tapetis* bacteria from Manila clam shells previously diagnosed with Brown Ring Disease. Our results reveal marine mollusc shells as novel genetic archives of the past, which opens new perspectives in ancient DNA research, with the potential to reconstruct the evolutionary history of molluscs, microbial communities, and pathogens in the face of environmental changes. Other future applications include conservation of endangered mollusc species and aquaculture management.

INTRODUCTION

Shells are biomineral exoskeletons offering protection from predation and desiccation to the soft bodies of invertebrates (Marin *et al.* 2012). Molluscs are prominent producers of shells, which despite microstructure variation among species, all consist of both a calcified and an organic layer (0.01-5.00% in weight) containing pigments, polysaccharides, lipids and proteins (Marin *et al.* 2012). Marine mollusc shells survive in the fossil record and provide invaluable information about the past. Their morphology can for instance be used for phylogenetic inference (Ponder & Lindberg 2008), while their microstructure can reveal important life history traits, including growth rates (Chauvaud *et al.* 2012) and diseases (Paillard *et al.* 2004; Trinkler *et al.* 2010a). Their internal structure is also commonly examined for sclerochronological dating (Gröcke & Gillikin 2008; Butler *et al.* 2013) or uncovering past changes in seawater temperature and salinity (Hiebenthal *et al.* 2012; Reynolds *et al.* 2016). Another source of information is their biogeochemical composition, with oxygen and carbon

Accepted Article

stable isotopes providing proxies for past temperatures (Chauvaud *et al.* 2005) and food availability (Lartaud *et al.* 2010); trace elements for the presence of pollutants (Pérez-Mayol *et al.* 2014) and phytoplankton dynamics (Szymanek 2011); and rates of amino-acid degradation for stratigraphic reconstruction (Penkman *et al.* 2011).

DNA extracted from marine mollusc shells can potentially offer an extremely informative and complementary proxy. As DNA from both inside and outside the shell could be entrapped during biomineral formation, shell DNA might be used not only for marine mollusc species identification, phylogeographic and demographic reconstructions, but also for bacterial community and environmental DNA profiling. Despite such potential, no work has evaluated the long-term preservation of DNA in marine mollusc shells, and hence, their potential for ancient DNA studies. To the best of our knowledge, only one single study has reported successful PCR amplification of nine nuclear microsatellites and a single mitochondrial (mtDNA) barcode from shells of freshwater pearl mussels *Margaritifera margaritifera* (Geist *et al.* 2008).

Genome-scale data from ancient specimens, including Middle Pleistocene remains (120-780,000 years Before Present, yBP) (Orlando *et al.* 2013; Dabney *et al.* 2013a; Meyer *et al.* 2014, 2016), reveal a wealth of information about past organisms, populations and environments (see (Pedersen *et al.* 2015; Ermini *et al.* 2015; Slatkin & Racimo 2016; Leonardi *et al.* 2016; MacHugh *et al.* 2016) for reviews). This has mainly been made possible through constant improvements in high-throughput DNA sequencing (HTS) technologies that have recently revolutionized genomics, metagenomics and transcriptomics (Bernstein *et al.* 2007; Metzker 2010; Segata *et al.* 2013; Crosetto *et al.* 2015; Orlando *et al.* 2015). Ancient DNA studies have so far mostly relied on the long-term preservation of DNA in teeth and bones (Smith *et al.* 2003), which are composed of calcium hydroxyapatite biominerals (Beniash 2011). Calcium carbonate matrices of sub-fossilized avian eggshells (Nys *et al.* 2004) have also been shown to occasionally preserve ancient DNA (Oskam *et al.* 2010). This work paves the way for the recovery of DNA from ancient marine mollusc shells, also made of calcium carbonate. To assess marine mollusc shell DNA

Accepted Article

content and investigate its long-term preservation, we extracted, shotgun-sequenced and characterised DNA from marine mollusc shells dated from 5 to ~7,000 yBP. We selected shells from clams (*Ruditapes philippinarum*, *Ruditapes decussatus*, *Venerupis corrugata*), abalones (*Haliotis tuberculata*), oysters (*Crassostrea sp.*), scallops (*Pecten maximus*), mussels (*Mytilus sp.*), and ocean quahogs (*Arctica islandica*) for the economic and ecological importance of these species, their common occurrence in European archaeological sites over long time ranges, and their variability in biomineral microstructure (Table 1). We applied both metagenomic and targeted mapping analyses with particular focus on molluscs, bacteria, chlorophytes and trematodes. This work establishes mollusc shells as (ancient) DNA reservoirs for studying (past) species, populations, and communities, as well as host/pathogen co-evolutionary arms races.

MATERIAL and METHODS

Shell sample-set description

We analysed clean marine mollusc shells, i.e., deprived of any soft tissue, from archaeological sites and collections (Tables 1 and S1). The specimens were dated between 5 and ~7,000 yBP based on their archaeological context (radiocarbon dates, ceramics, lithic artefacts (Large & Birocheau 2004; Andersen 2008; Gissinger 2011; Champagne *et al.* 2012; Le Goff & Dupont 2015)) or collection records. The sample hav_Mussel was radiocarbon-dated as part of this study at the Kech Carbon Cycle AMS Facility, University California Irvine, leading to a 95.4% probability calibrated range of 6,809-6,962 yBP (UCIAMS#182492, Marine13 calibration curve (Reimer *et al.* 2013)). Diagnostic macroscopic morphological characters were used for identifying species. A number of shells exhibited damage potentially indicative of microbial or trematode infection (Tables 1 and S1). More specifically, *R. philippinarum* Manila clam shells presented growth lesions and dark conchiolin deposits in their inner surface that are typical of the Brown Ring Disease (BRD) caused by the bacterium *Vibrio tapetis* (Paillard *et al.* 1989;

Accepted Article

Paillard & Maes 1995). Other clam shells pre-dating the first characterisation of BRD (Paillard *et al.* 1989) also exhibited weak conchiolin deposits.

Shell sample preparation

For shells dating from the last 60 years (and hereafter referred to as "modern"), DNA extraction was performed on a fragment of the shells' ventral margin. For older shells ("ancient"), either a fragment or the whole shell was used.

We first assessed the feasibility of recovering DNA from the "modern" marine mollusc shells by analysing both a healthy (NEG_Clam) and a BRD-positive (POS_Clam) Manila clam (~27 yBP; Table 1, Figure S1). Based on positive preliminary results, the impact of shell pre-extraction bleach decontamination was investigated by comparing shotgun HTS datasets for bleach-treated (KOR_I^{-}B^{+}_Clam, KOR_I^{+}B^{+}_Clam) and non-treated (KOR_I^{-}B^{-}_Clam, KOR_I^{+}B^{+}_Clam) shell fragments of a BRD-positive Manila clam (KOR_Clam) (Table 1, Figure S1). DNA was analysed both from "diseased" (KOR_I^{+}B^{-}_Clam, KOR_I^{+}B^{+}_Clam) and "healthy" (KOR_I^{-}B^{-}_Clam, KOR_I^{-}B^{+}_Clam) fragments of the same shell. Subsequent to this test, we decontaminated the shells/shell fragments through incubation in one 1% bleach volume for 10 min under constant mixing, followed by three washes in three volumes of distilled water (Figure S1). After drying, the shells were ground to fine powder using a pestle and mortar.

DNA extraction

DNA was extracted from the shell powder based on the so-called "Y1" method described in (Yang *et al.* 1998; Gamba *et al.* 2014, 2016). In short, powder was incubated overnight at 37°C in 15 mL digestion buffer (0.45 mM EDTA, 0.5% N-laurylsarcosyl, 0.25 mg/mL proteinase K) under constant mixing. After centrifugation (2,000 RPM; 2 min), the supernatant was filtered

Accepted Article

through an Amicon Ultra-15 30kDa centrifugal filter unit (Merck Millipore) (3,000 RPM; 50 min). DNA was then purified using the MinElute PCR Purification Kit (QIAGEN) in a final elution volume of 60 µL EB buffer.

Library construction and sequencing

Extracts were built into blunt-ended DNA libraries for Illumina shotgun sequencing based on the protocol described in Seguin-Orlando *et al.* (2013) and Orlando *et al.* (2013). We used the NEBNext Quick DNA Library Prep Master Mix Set for 454 (New England BioLabs) in 50 µL for end-repair and ligation (0.5 µM Illumina adapter final concentration) and in 25 µL for fill-in. End-repair was carried out at 12°C for 20 min and at 37°C for 15 min, ligation at 20°C for 20 min, fill-in at 37°C for 20 min followed by *Bst* enzyme inactivation at 80°C for 20 min. We used the MinElute kit for purification after end-repair and ligation with EB buffer elution volumes of 30 µL and 20 µL, respectively.

Prior to Illumina shotgun sequencing, the libraries were amplified by PCR. The number of cycles required to recover DNA library amounts sufficient for sequencing, while minimizing library clonality, was estimated on a per-sample basis by qPCR carried out in duplicates in 20 µL of the following reaction mix: 1 µL 1/20 diluted DNA library (in EB buffer), 0.1 units Taq Gold, 1X Gold Buffer, 4 mM $MgCl_2$ (Life Technologies), 1 mg/mL BSA, 0.25 mM of each dNTP, 0.2 µM of Primer inPE1.0 (5'-AATGATACGGCGACCACCGAGATCTAC-ACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and 0.2 µM of an Illumina 6 bp-indexed ('I') primer (5'-CAAGCAGAAGACGGCATACGAGATIIIIIIGTGACTGGAGTT-CAGACGTGTGCTCTTCCG-3'), and 0.8 µL of 1:4:2000 ROX:SybR:DMSO DNA dye mix (ThermoFisher Scientific). Thermo-cycling conditions were: activation at 92°C for 10 min; 40 cycles of: denaturation at 92°C for 30 sec, annealing at 60°C for 30 sec, elongation at 72°C for 30 sec; final elongation at 72°C for 7 min. qPCRs were performed on a LightCycler 480 Real-Time PCR System instrument (Roche

Accepted Article

Applied Science). A relative measure of DNA concentration for each library was provided by the Ct value estimated using the second derivative maximum method with the LightCycler Software 4.0. DNA contamination from the laboratory reagents/environment was controlled by analysing blanks at the same time as samples during extraction (Extraction Blank Control, EBC), library building, and qPCR. Extraction success rates were calculated considering an extraction successful when a library's qPCR Ct value was inferior to the Ct value minus 7 of the EBC from the corresponding extraction session (≥128-fold concentration compared to EBCs). Only modern shell DNA libraries fulfilling this condition were selected for shotgun sequencing. As this was not satisfied for most ancient shell DNA libraries, those obtained from their corresponding EBCs were independently PCR-amplified and indexed with unique external barcodes, before being pooled and sequenced with the amplified ancient shell DNA libraries (Table S1).

Each library was PCR-amplified using one identifying indexed Illumina primer (Table S1). Library amplification was performed in the same conditions as for qPCR, except that 5 µL of undiluted DNA extract was added to 20 µL of the qPCR reaction mix without ROX:SybR:DMSO, and that the number of cycles was derived from Ct values for each library (6-21 for shell DNA libraries, 15-29 for EBC libraries; Table S1). After MinElute purification and elution in 25 µL EB buffer, the amount of DNA in the amplified indexed libraries was quantified on a 2200 TapeStation instrument (High Sensitivity D1000 ScreenTape, Agilent). DNA amounts in blanks (except EBCs amplified for sequencing) were below the detection threshold of the instrument, in line with low levels of DNA contamination during laboratory work. The indexed DNA libraries were pooled and sequenced in either paired-end or single-end mode on the Illumina HiSeq2000 and MiSeq platforms at the Danish National High-Throughput DNA Sequencing Centre (Table S1).

Accepted Article

DNA sequence read processing and mapping

Post-sequencing read processing and mapping were carried out using PALEOMIX (Schubert *et al.* 2014), including adapter trimming with AdapterRemoval2 (Schubert *et al.* 2016), collapse of overlapping pair-end mates as described in (Orlando *et al.* 2013), and read mapping against reference DNA sequences (see below and Tables S2-S6) with BWA v0.5.9 (Li & Durbin 2009). Seeding was disabled for mapping, as recommended in (Schubert *et al.* 2012), and only reads showing mapping qualities ≥30 were kept. PCR duplicates were removed with MarkDuplicates in Picard Tools version 1.119 (http://broadinstitute.github.io/picard/) for non-collapsed reads, and the PALEOMIX FilterUniqueBAM Python script (Schubert *et al.* 2014) for collapsed reads.

Metagenomic analyses using DIAMOND

To characterise the total DNA content of marine mollusc shells, we used the metagenomic program DIAMOND v0.8.17 (Buchfink *et al.* 2015) and the BLAST "*nr*" protein database (downloaded on the 27/07/2016) with default parameters and the "--sensitive" alignment mode (Figure S1). For POS_Clam and NEG_Clam, non-collapsed paired-end reads were disregarded in order to limit the impact of potential contaminating modern DNA, building on the expectation that these would be longer than damaged endogenous DNA molecules. Taxonomic classification was then performed by the program daa-meganizer distributed with MEGAN6 (Huson *et al.* 2007, 2011) using the weighted lowest-common algorithm (option "-wlca") and only keeping the top-1% hits ("-top 1.0") showing sequences identical to database markers ("-mpi 100.0") (Table S7).

Accepted Article

Mollusc DNA reference sequences

The presence of mollusc nuclear DNA was examined by mapping shotgun HTS reads to the genome sequences published for the Pacific oyster *Crassostrea gigas* (Zhang *et al.* 2012) and the blue mussel *Mytilus galloprovincialis* (Murgarella *et al.* 2016), as well as to sets of transcripts for the abalone *H. tuberculata* (Harney *et al.* 2016), the scallop *P. maximus* (Pauletto *et al.* 2014; Artigaud *et al.* 2014, 2015) and *M. galloprovincialis* (Moreira *et al.* 2015) (Figure S1, Table S2). For the clam *R. philippinarum*, transcript sequences were collected from the Ruphibase database (http://compgen.bio.unipd.it/ruphibase/), the NCBI EST database (https://www.ncbi.nlm.nih.gov/nucest/) and previous publications (Milan *et al.* 2011; Moreira *et al.* 2012a; b; Ghiselli *et al.* 2012; Allam *et al.* 2014) (Table S3).

Mollusc mtDNA content was estimated by mapping shotgun HTS reads to two reference sequence databases. The first database contained 52 mtDNA genomes for the mollusc species represented in our sample-set, or closely related species (Figure S1, Table S2). To account for the circularity of mtDNA genomes, the first 30 bp in each reference sequence were manually duplicated and added at the end of the corresponding sequence as in (Vilstrup *et al.* 2013). For each sample, DNA reads were mapped to each reference sequence independently in order to control for false positive species identification. The second mtDNA database contained sequences of the *cytochrome c oxidase subunit I* gene 5'-extremity (COI5P), for which 88,453 barcodes are available for the Mollusca phylum in the Barcode Of Life Database (http://www.boldsystems.org/) (Figure S1). Hits to the database were reported as the total read length mapping to all the markers for a given taxon; in cases where this length was <100bp, the corresponding taxa were excluded.

Vibrio DNA reference sequences

We next assessed the potential of shell DNA for the molecular identification of *V. tapetis* in BRD-positive shells. We mapped shotgun HTS reads to genomic sequences of two *V. tapetis* strains, RP2-3 and HH6087, generated as part of this study (Figure S1, Table S4) and isolated

Accepted Article

respectively from a BRD-positive *R. philipinarum* clam in France in 1990 (Paillard *et al.* 1994; Borrego *et al.* 1996), and from a healthy Atlantic halibut (*Hippoglossus hippoglossus*) in Scotland in 2002 (Reid *et al.* 2003). DNA was extracted following (Bidault *et al.* 2015). Shotgun HTS reads were obtained by applying the protocol described above on 1 µg of DNA extract sheared using a diagenode Bioruptor ultrasonicator (four cycles of: 15 sec at high intensity; 90 sec off), and amplifying libraries for seven cycles before single-end sequencing on an Illumina HiSeq2000 platform at the Danish National High-Throughput DNA Sequencing Centre. Post-sequencing, the adapters were removed using AdapterRemoval v2.1.5 (Schubert *et al.* 2016). De novo assembly was performed using the SPAdes assembler v3.7.0 (Bankevich *et al.* 2012), using four k-mer sizes (21, 33, 55 and 77), and allowing for mismatch correction. Only contigs ≥300 bp in length were kept for further analyses.

To verify that shell HTS reads mapping to the *V. tapetis* reference sequences belonged to the *Vibrio* genus, they were aligned to the BLAST "*nt*" nucleotide database (downloaded on 11/10/2016) using the BLAST2 aligner (Altschul *et al.* 1990; Camacho *et al.* 2009), followed by MEGAN6 analyses as described above.

To rule out potential false positive identification of *Vibrio* species, shell HTS shotgun reads were also mapped to 63 genomic sequences corresponding to one representative sequence per *Vibrio* species so far characterised at the genome level (Table S4). The obtained mapping patterns were compared to those expected when mapping reads deriving from each *Vibrio* genome against every other *Vibrio* reference sequence. This was done by simulating 101 bp HiSeq2000 single-stranded reads from each *Vibrio* genomic sequence using ART (version Mount-Rainier (Huang *et al.* 2012)). The number of reads simulated for each *Vibrio* sequence corresponded to a ~7X average depth-of-coverage, the maximal coverage obtained when mapping shell HTS DNA reads to *V. tapetis* reference sequences.

Accepted Article

Chlorophyte, trematode and human DNA reference sequences

We attempted to confirm the results of the DIAMOND metagenomic analyses identifying Chlorophyta green algae in shells, by mapping shotgun HTS reads to *(i)* the complete mtDNA and chloroplast genomes of *Pseudendoclonium akinetum* (Genbank accession numbers AY359242 and AY835431), *(ii)* the complete chloroplast genomes of *Bryopsis plumosa*, *Phacotus lenticularis* (LN810504, KT625422), and; *(iii)* the complete plastid genome of *Chlorella variabilis* (HQ914635) (Table S5). We also investigated the presence of trematode DNA in shells through HTS read mapping to the complete mtDNA sequences of the candidate trematodes identified by DIAMOND (with the exception of *Opisthorchis viverrini* for which no comparative data are available): *Clonorchis sinensis* (FJ381664), *Schistosoma mansoni* (HE601612), *S. haematobium* (DQ157222), *S. japonicum* (HM120841), *S. mekongi* (NC002529), and *S. spindale* (DQ157223) (Table S6). Human contamination was estimated by mapping HTS reads to the *Homo sapiens* complete nuclear and mtDNA reference genomes (assembly hg19 build 37.1).

Post-mortem DNA damage profiles

DNA fragmentation, base misincorporation, and fragment size distributions in shell DNA were assessed using mapDamage v2.0.1 (Jónsson *et al.* 2013) and 100,000 MCMC iterations. Damage was investigated for mollusc, *V. tapetis*, chlorophyte and trematode DNA, considering both all reads and randomly sampling from bam files the minimum number of sequencing reads mapped across samples for a given marker. For each marker and each sample yielding an accurate taxonomic assignment, we used as input file the bam alignment for which the maximal coverage was obtained (Tables S8-S10). We estimated the following damage parameters: rates of deamination in double strands (δ_D) and single strands (δ_S), probability of reads not terminating in overhangs (λ, transformed into $1/\lambda-1$, a proxy for the length of overhanging regions). We then contrasted nuclear and mtDNA damage parameters for each single ancient

Accepted Article

sample against modern samples. We compared the proportion of mapDamage iterations belonging to the damage parameter's 97.5% confidence interval in ancient samples that overlapped the 97.5% confidence interval of modern samples, considering each mollusc species individually or the whole set of species altogether. This number was reported as the percentage of overlap between the damage parameter 97.5% confidence interval for the considered ancient sample and modern samples (Table S11).

Microbial profiling using metaBIT

The microbial content of shell DNA extracts was characterised using metaBIT (Louvel *et al.* 2016) (Table S7). As in MEGAN analyses, non-collapsed paired-end reads were disregarded for POS_Clam and NEG_Clam. Microbial taxon assignment and relative abundance calculation were carried out by first aligning shotgun reads to the MetaPhlAn2 database (Segata *et al.* 2012; Truong *et al.* 2015) using the Bowtie2 v2.1.0 aligner (Langmead & Salzberg 2012). PCR duplicates were then removed as described above.

For comparison purposes, shotgun data from 60 seawater samples (deep chlorophyll maximum, N=20; mesopelagic layer, N=14; surface, N=26; Table S12), all published by the Tara Oceans Expedition (Sunagawa *et al.* 2015), were profiled using metaBIT as detailed above, except that all reads, including paired-reads, were considered in the analyses. We also used the comparative microbial profiles distributed with metaBIT including 689 human (Human Microbiome Project Consortium 2012) and 15 soil samples (Fierer *et al.* 2012; Louvel *et al.* 2016).

Statistical analyses of the MetaPhlAn2 profiles (Truong *et al.* 2015) were performed in metaBIT, only considering taxa showing abundances ≥1% in order to reduce noise due to false-positives.

Accepted Article

RESULTS

Preliminary analyses and impact of bleach decontamination

DNA amounts compatible with shotgun HTS could be extracted from the 27 year-old shells of the Manila clams NEG_Clam and POS_Clam. For both samples, metagenomic comparison of the shotgun HTS reads with the BLAST *nr* database using DIAMOND and MEGAN6 revealed that most reads were assigned to the Bacteria kingdom (Figure 1A). A fraction (1.3-4.0%) of these bacterial reads was assigned to the *Vibrio* genus for both shells, and to the *V. tapetis* species for the BRD POS_Clam only (2.4%; Figure 1A). HTS reads mapped to the genome sequences of the two *V. tapetis* strains HH6087 and RP2-3 with a 29-369-fold enrichment (after sequencing effort normalization) in POS_Clam compared to NEG_Clam (Figure 1B). This suggests that DNA assigned as *V. tapetis* can be retrieved from BRD-infected marine mollusc shells, that it does not arise from contamination, and that other *Vibrio* species might represent members of the healthy shell microbiota (Lokmer & Wegner 2015; Lokmer *et al.* 2016). Within Mollusca, the most assigned eukaryotic phylum (Figure 1A), we identified the expected *R. philippinarum* species by mapping to transcript (0.4-2.0X maximal coverage) and mtDNA reference sequences (3.1-29.0X) (Figure 1B, Table S8). The datasets contained minimal fractions of DNA arising from contamination during collection, analyses, curation, and/or sampling, as 0.17% (NEG_Clam) and 0.12% (POS_Clam) of the HTS reads (Figure 1B) mapped to the human reference sequence.

Bleach decontamination prior to DNA extraction of KOR_Clam reduced the fraction of reads mapping to the human genome (0.03% in bleached *vs* 0.05-0.49% in non-bleached shells; Figure 1B) and being assigned to the Bacteria kingdom (Figure 1A). This, however, preserved mollusc DNA content (Figure 1B) with a sum of 1.39-3.82% of the reads mapping to the *R. philippinarum* transcripts and mtDNA genomes in bleached shells *vs* 2.68-2.96% in non-bleached shells. This was true even after correcting for the amount of shell powder used (0.05-0.09% in bleached *vs* 0.07-0.13% in non-bleached samples). For the two BRD-positive fragments, bleach treatment did not result in a loss of reads mapping to the genome of *V. tapetis* RP2-3 (3.22% in bleached *vs*

Accepted Article

1.83% in non-bleached infected shells; both 0.06% when correcting for the amount of shell powder used for extraction; Figure 1B). Finally, bleach treatment did not increase DNA damage levels (estimated rates of deamination and overhang length) in mollusc and *V. tapetis* DNA (Figure S2). As bleach treatment only impacted exogenous DNA sources and maintained host and pathogenic DNA content, we applied pre-extraction bleach decontamination in all subsequent analyses.

DNA extraction success rate

DNA extraction success rates were variable across sample age and species. In modern samples, we could successfully extract DNA from 16/21 shells (76%), which included all those for clams (11/11) and abalones (3/3), but a limited number of shells for scallops (2/5) and one diseased oyster (1/3). Ancient samples yielded a lower success rate of 5/17 (29%), with successful extraction for 2/10 clams, none of the 2 mussels, 2/2 abalones and 1/3 ocean quahogs.

Metagenomic content

Metagenomic analyses revealed that the most abundant kingdom in the majority of HTS datasets was Bacteria followed by Eukaryota, Archaea and viruses (Figure 2A). Mollusca was the most abundant eukaryotic phylum (Figure 2B), except for two samples showing highest abundances for Chlorophyta green algae, Phaeophyceae brown algae, Bacillariophyta diatoms, Basidiomycota fungi and Nematoda. Importantly, Mollusca were virtually absent from all EBCs, suggesting no contamination from exogenous mollusc DNA. Within Mollusca, success rates were variable when attempting to identify, based on maximum number of hits per sample, morphologically-defined mollusc genera (Figure 2C): 20-100% for modern and 0-100% for ancient shells (Table 2) when considering the totality of samples extracted, including those not sequenced. All misidentified specimens, but no EBCs, were assigned to *Mytilus*, suggesting

misidentification likely due to the over-representation of *Mytilus* markers in the *nr* database (i.e., 3,090 *vs* 428 for *Ruditapes,* 21 for *Venerupis*, and 43 for *Arctica*) rather than to contamination. Within Chlorophyta, only POS_Clam and BRE5m_Oyster led to taxonomic identification at the genus level (*Pseudendoclonium, Ostreobium*, *Chlorella*, *Bryopsis*, *Phacotus, Derbesia*, and *Caulerpa*; Figure 2D). We also investigated the presence of DNA from trematodes as possible parasites, and found, in fifteen shells, high-quality hits against *Schistosoma mansoni*, *Clonorchis sinensis* and *Opisthorcis viverrini* (Figure 2E).

Microbial profiling

We further characterised the microbial content of shell HTS datasets using metaBIT. Principal Coordinate Analysis (PCoA) of Bray-Curtis distances at the genus level showed that the microbial diversity recovered from the marine mollusc shells fell outside the human microbiome diversity (Figures 3A and S3A). Two samples, hel_Quahog and NEG_Clam, however showed affinities for the human skin/nose microbiomes, due to elevated abundances in the skin-associated bacteria *Propionibacterium acnes* (42.8% and 54.6%), suggesting human-derived contamination in these samples that also showed maximal levels of human DNA (0.21% and 0.62%) (Figure S4). Although this result can be explained by the absence of pre-extraction decontamination for NEG_Clam, it also exemplifies the limit of our bleach decontamination procedure for hel_Quahog.

Marine mollusc shell microbial profiles showed most affinity with those of soils, and segregated from the seawater microbial diversity (Figures 3B and S3B), probably not representative of the molluscs' mostly coastal environment. metaBIT microbial profiling most likely suffers from high false negative rates due to the limited diversity of environmental microbe markers present in the MetaPhlAn database (Louvel *et al.* 2016). Although bacterial species typical of marine seawaters and sediments could be identified (Figure S5), our results call for a thorough shotgun HTS-based characterisation of the microbiomes in mollusc tissues (Desriac *et al.* 2014; Lokmer

Accepted Article

Accepted Article

& Wegner 2015; Meisterhans *et al.* 2015; Lokmer *et al.* 2016) and in coastal environments (Mustafa *et al.* 2016). This would improve the resolution power in profiling shell microbial sources, and help understand the interplay between molluscs and climatic changes, pollutants and emerging pathogens.

Molecular identification of molluscs

After mapping shell shotgun reads against a comparative sequence database for molluscs, taxonomic identification was considered positive when maximal coverage was obtained for a reference sequence representing the expected genus for a given sample.

Based on nuclear data and mtDNA genomes, most modern samples could be taxonomically assigned to the right genus, and sometimes even species, (Figures 4A-B, Tables 2 and S8) with depths-of-coverage of 0.006-2.008X and 2.057-54.142X, respectively. Standing as exception, MURp_Clam was misassigned to *Mytilus trossulus* based on mtDNA genome mapping (0.057X maximal coverage), as was LEF-B_Scallop (0.010X) also mis-assigned to *H. tuberculata* based on nuclear data (0.003X). Both samples might show low mollusc DNA content, probably leading to spurious mapping also observed in EBCs. Genus identification using mtDNA COI5P barcodes was also robust, with genus assignment success for 10/10 clams, 3/3 abalones, 1/1 oyster and 1/2 scallops (Figure 4C, Table 2). Of note, MuRp_Clam, initially labelled as *R. philippinarum*, was identified as *R. decussatus* based on COI5P barcodes, which is in line with the pattern of coverage observed when mapping against mtDNA genomes in absence of reference sequences for *R. decussatus*. Reanalysis of the shell morphology did not dismiss the identification of MURp_Clam as *R. decussatus*.

Mollusc taxonomic identification success was more variable for ancient shells. Considering nuclear reference sequences (Figure 4A, Tables 2 and S8), none could be characterised at a coverage ≥0.1X for any ancient sample, except med_Abalone (0.429X). For ten samples,

Accepted Article

however, maximal coverage (0.001-0.039X) was obtained for the correct genus. Spurious mapping to the *P. maximus* (0.003X) and *H. tuberculata* (<0.001X) transcripts was observed in EBCs, as well as in samples with low endogenous mollusc DNA content (gripB_Clam and lmc2B_Clam; 0.001X), and/or for which comparative nuclear data from a closely related species was not available (hel_Quahog, quahog2_Quahog, and ves_Quahog; 0.001-0.004X). Mapping shotgun reads to mtDNA genomes, mollusc taxa could be positively identified for all ancient samples (0.223-6.665X coverage), except clams. For hav_Mussel, maximal coverage was obtained for the complete mtDNA genome of *M. galloprovincialis* AY497292 (0.223X), but relatively high coverage (0.145-0.215X) was also recovered for *M. trossulus* (DQ198231, KM192124-28, KM192130-32), *Mytilus edulis* (AY484747) and *M. galloprovincialis* (FJ890849, DQ399833), which all form a monophyletic clade of Baltic Sea taxa (Zbawicka *et al.* 2014). The mis-assigned shells were the *R. decussatus* and *V. corrugata* clams identified as *M. trossulus* or *M. edulis* due to spurious mapping (0.007-0.047X), similar to that observed in EBCs (0.002-0.007X). This could be explained by low endogenous mollusc content and/or by the lack of comparative mtDNA genome sequences for these species (Figure 4B, Tables 2 and S8). Except for hav_Mussel, positive genus identification using mitochondrial COI5P barcodes was successful for all ancient samples, when considering *Ruditapes* and *Venerupis* as closely related genera (Figure 4C, Tables 2 and S8).

Molecular identification of Vibrio tapetis in shells of infected Manila clams

We next aimed at identifying the *V. tapetis* species in BRD-positive Manila clam shells, using read alignment against the currently known diversity of *Vibrio* genomes (Figures 5 and S6, Table S9). Maximal coverage across all *Vibrio* species was obtained for the *V. tapetis* strain RP2-3 in the shells showing signs of acute BRD: POS_Clam (7.322 *vs* 1.932X for *V. tapetis* strain HH6087) and KOR_Clam (4.195 *vs* 1.095X). In Manila clam shells pre-dating the first description of BRD and showing signs of infection, maximal coverage was obtained for the other *V. tapetis*

Accepted Article

strain HH6087: LAN1p_Clam (1.910X *vs* 0.447X for *V. tapetis* strain RP2-3), LAN2p_Clam (0.672X *vs* 0.160X), and LAN3p_Clam (0.510X *vs* 0.122X). After normalizing by the sequencing effort, depths-of-coverage in infected samples were 145-1,897-fold higher than those in EBCs, suggesting a limited impact of contamination in our dataset (Figure 5 and S6, Table S9).

For POS_Clam, KOR_Clam, LAN1p_Clam, LAN2p_Clam and LAN3p_Clam, 59.5-72.1% of the reads mapping to the two *V. tapetis* reference sequences were assigned to the *Vibrio* genus after alignment to the BLAST *nt* database (Figure S7). For these samples, the *Vibrio* genome coverage was at least 12.6-45.8-fold higher for the *V. tapetis* species than any other *Vibrio* species presently described at the genome level, as expected when mapping *V. tapetis* reads to all other *Vibrio* reference sequences (Figures S8 and S9). These results are in line with the *Vibrio* DNA recovered from POS_Clam, KOR_Clam, LAN1p_Clam, LAN2p_Clam, and LAN3p_Clam belonging to the *V. tapetis* species or to a close relative hitherto not sequenced.

In shells showing signs of weak BRD infection, the depths-of-coverage for the two *V. tapetis* genomes were similar to those for other *Vibrio* species: 0.009X for MATp_Clam (<0.001-0.030X in other *Vibrio* species), <0.001X for MURp_Clam (<0.001-0.013X), and 0.004-0.005X for SELp_Clam (<0.001-0.052X). This suggests infection by another pathogen or a reduction in *V. tapetis* DNA content below our analyses' detection threshold (Figures 5 and S6). Likewise, shells showing conchiolin deposits different from those observed in BRD Manila clams (AKKp_Clam, BRE3m_Abalone, BRE3p_Abalone, BRE5m_Oyster, BRE5p_Oyster) did not show specific enrichment for *V. tapetis* (Table S9, Figures 5 and S6). In healthy med_Abalone and BRE2_Abalone, *V. tapetis* HH6087 showed maximal coverage (0.06X and 0.11X, respectively), which was too low to allow robust genome reconstruction (Figure 5 and S6, Table S9). In the abnormal AKKp_Clam shell, maximal coverage was obtained for *V. rotiferianus* (0.263X *vs* 0.001-0.177X in other *Vibrio* species), a known pathogen of marine organisms (Chowdhury *et al.* 2011) (Figure S6, Table S9).

Accepted Article

Post-mortem DNA damage patterns

We next characterised the preservation of DNA molecules recovered from shells, which also provided evidence for our data authenticity.

First, when mapping shotgun reads from ancient shells against mollusc reference sequences, we observed an increase (decrease) of purines (pyrimidines) at genomic positions just before read starts due to depurination in the mollusc nuclear DNA of some ancient and modern shells (Figures 6 and S10).

Second, we observed C→T and G→A misincorporation patterns due to cytosine deamination typical of ancient DNA (Dabney *et al.* 2013b) (Figures 6 and S11, Table S10). C→T substitution rates at the 5'-end of ancient nuclear mollusc DNA fragments varied from 2.2% in the 106 year-old thi_Clam to 16.8% in the ~5,000 year-old auzay1B_Clam. In modern samples, these rates varied from 1.4% in the 27 year-old NEG_Clam to 3.0% in the 10 year-old BRE5m_Oyster. In clams, where sample sizes allow the comparison, 5'-C→T rates were significantly higher than those observed in modern shells (median(ancient)=10.1%; median(modern)=1.8%; Wilcoxon test *p*-value= $4.57\ 10^{-5}$) (Figures 6, S11 and S12, Table S10).

Some ancient samples showed significant differences with modern samples when considering mapDamage estimates of deamination in double-strands (δ_D) and single-strands (δ_S), and of the length of overhanging ends ($1/\lambda$ - 1). For nuclear mollusc DNA, the range of overlap between the 97.5% confidence interval for all modern samples and the 97.5% confidence interval for each ancient shell was 6.4-99.4% for δ_D, 9.7-97.7% for δ_S, and 12.4-97.9% for $1/\lambda$ – 1 (Table S11; Figure S13). For mollusc mtDNA, this overlap was increased with ranges of 62.9-97.7% for δ_D, 63.1-98.8% for δ_S, 47.8-99.3% for $1/\lambda$ – 1 (Table S11; Figure S14). For clams and abalones, where damage parameter distributions for ancient and modern samples of the same species can be compared, we obtained overlap values similar to those calculated considering all modern samples (Table S11). Compared to δ_D, δ_S was 4.1-35.3-fold higher in ancient mollusc nuclear

Accepted Article

DNA and 10.9-107.3-fold higher in ancient mollusc mtDNA (Table S10). These damage-driven patterns are characteristic of ancient DNA and reflect the variability in preservation of DNA molecules recovered from shells.

Nucleotide misincorporation typical of ancient DNA was also present in the *V. tapetis* DNA recovered from 12-32 year-old BRD-positive Manila Clam shells, with C→T rates at 5'-ends varying from 0.8% in POS_Clam to 1.7% in LAN1p_Clam and LAN2p_Clam (Figures 6 and S15). In *V. tapetis* DNA, δ_S estimates were 9.8-42.2 fold higher than δ_D, as expected (Table S10 and Figure S16).

Third, we observed various levels of damage-driven DNA fragmentation across the dataset. Ancient shells showed high fragmentation with average 43-50 bp mollusc DNA fragment sizes (hav_Mussel, auzay1B_Clam, auzay2B_Clam, auzay3B_Clam, auzay4B_Clam, quahog2_Quahog, ves_Quahog), intermediate fragmentation (lmc1B_Clam, lmc3B_Clam med_Abalone, nis_Clam, ves_Quahog, thi_Clam) and long fragment sizes (gue_Abalone) (Figures 6 and S17). In mollusc nuclear DNA from modern shells, low fragmentation was observed in LEF-B_Scallop, ROS-B_Scallop, BRE1_Abalone, BRE3m_Abalone, KOR_Clam, BRE2_Abalone, BRE5m_Oyster, and BRE5p_Oyster (Figures 6 and S18). Interestingly, mollusc nuclear DNA fragment size distributions in LAN1p_Clam, LAN2p_Clam, LAN3p_Clam, SELp_Clam, BRE3p_Clam, AKKp_Clam and MATp_Clam showed a ~10 bp periodicity previously proposed to reflect nucleosome protection (Pedersen *et al.* 2014) (Figures 6 and S18). This is supported by this periodicity being absent from mollusc mtDNA and *V. tapetis* DNA (Figures 6, S15 and S19), which are not associated with histones.

Molecular detection of trematodes

We could not robustly confirm the DIAMOND-based metagenomic identification of trematodes by mapping shell shotgun HTS reads against complete mtDNA sequences. Only one sample, LAN1p_Clam, showed coverage above 0.10X (0.14X to the *S. mekongi* reference sequence), but

Accepted Article

no DNA damage signal. No reads from nis_Clam mapped against the *S. mansoni* reference sequence, despite the fact that the shell exhibited perforation lesions proposed to be caused by trematodes, and that *S. mansoni* was detected in metagenomic analyses. For BRE5m_Oyster and BRE5p_Oyster showing signs compatible with trematode attack, no reads mapped to any trematode reference sequence.

Molecular detection of chlorophytes

The putative presence of chlorophytes indicated by DIAMOND was supported only marginally for *C. variabilis* and *B. plumosa* when mapping shell shotgun HTS reads to their complete plastid and chloroplast reference genome sequences, respectively. For *C. variabilis*, a slight enrichment (after normalizing for sequencing effort) was observed compared to EBCs (<0.001-0.039X) for POS_Clam (0.707X; 4.6-fold enrichment), BRE5m_Oyster (0.270X; 5.1-fold enrichment), MATp_Clam (0.170X; 2.0-fold enrichment), hel_Quahog (0.124X; 2.1-fold enrichment), and AKKp_Clam (0.090X; 2.0-fold enrichment). For *B. plumosa*, POS_Clam (0.486X; 4.2-fold enrichment), BRE5m_Oyster (0.427X; 10.8-fold enrichment), and hel_Quahog (0.09X; 2.1-fold enrichment) showed higher depths-of-coverage than in EBCs (<0.001-0.027X). The low number of unique high-quality DNA reads mapping to these chlorophyte reference sequences was not compatible with robust identification of damage authentication signatures in mapDamage.

DISCUSSION

Long-term DNA preservation in marine mollusc shell biominerals

Here we demonstrate that authentic endogenous DNA can be extracted from the shell of marine molluscs, including ancient specimens covering the last ~7,000 years. The mollusc species investigated did not show the same DNA extraction success rates, owing to their different

Accepted Article

biomineral microstructures, preservation, and disease state. DNA extraction was more successful for clams, abalones and ocean quahogs, with a shell inner layer composed of homogeneous or columnar nacreous aragonite, than oysters and scallops, both characterised by a foliated calcite inner shell microstructure. Although the presence of nucleic acids in mollusc shells has not been documented to date, differences in DNA content could be due to differences in shell organic content, as nacre and prism microstructures contain a larger organic component (~1% in weight) than crossed-lamellar foliated microstructures (Marin *et al.* 2012). Additionally, the denser structure of nacreous/prismatic matrices compared to calcite might offer shell DNA a better protection against post-mortem hydrolysis or microbial attack. Interestingly, BRE5_Oyster, which showed an important brown deposit on the inner layer indicative of a diseased state, was the only oyster providing positive results. We hypothesize that the mollusc defence mechanisms, which involve increased mineralisation and cell density and/or the recruitment of hemocytes in the hemolymph and extra-pallial fluid (Paillard *et al.* 1996), may have increased DNA content in this particular shell.

We observed that DNA extraction success rates from ancient marine mollusc shells depended on the preservation of the inner layer. Macroscopic examination of clam shells dated ~200-5,000 yBP and yielding low amounts of DNA, revealed a loss of the "porcelaneous" aspect of the inner layer due to the dissolution of aragonite during deposition. We therefore recommend that particular attention be paid to the integrity of the inner layer (preferably aragonitic) when selecting samples for ancient DNA analyses. Further investigation is required to understand the interactions between DNA, proteins and biominerals, especially regarding biomolecular taphonomy in calcium carbonates, and to assess temporal ranges compatible with shell DNA recovery.

Long-term DNA survival within shell calcium carbonate biominerals was supported by degradation patterns characteristic of ancient DNA. These, including increased rates of C→T and G→A mutations at read termini and ~10-bp periodicity in the size distribution of nuclear

Accepted Article

DNA fragments, also suggests potential for future epigenetics studies in molluscs, as methylation and nucleosome maps can now be reconstructed from patterns of nucleotide misincorporation and depths-of-coverage (Pedersen *et al.* 2014; Gokhman *et al.* 2014; Orlando & Willerslev 2014; Orlando *et al.* 2015; Hanghøj *et al.* 2016).

Metabarcoding of marine mollusc shells

Among the different computational methods used here, the metagenomic approach based on DIAMOND and MEGAN proved useful to screen for possible candidate taxa. Further confirmation for the presence of chlorophyte, trematode and bacterial DNA would require deeper sequencing or whole-genome in-solution target enrichment for specific DNA markers (Bos *et al.* 2011; Schroeder *et al.* 2015; Ávila-Arcos *et al.* 2015; Cruz-Dávalos *et al.* 2016). The metagenomic approach is however computationally heavy and subject to uneven taxon representativeness in comparative databases which, in addition to the absence of normalization for marker sizes, biases the relative taxon abundance estimates, possibly leading to erroneous taxon identification. For that purpose, more specificity and accuracy were achieved when mapping to target reference sequences. This approach however requires prior knowledge regarding mollusc genera/species to target, which is not always available, e.g., in cases where shell samples are too fragmentary for robust morphological identification. Targeted mapping is also limited by the amount of data being publicly available, e.g., here, for *A. islandica* and *R. decussatus*. Nevertheless, with the increasing application of HTS to non-model organisms, we can be confident that more comparative mollusc datasets will be made available soon. We also show here that mollusc mtDNA barcode sequences can be recovered from shells, making PCR-based metabarcoding applicable to marine mollusc shell DNA (Geist *et al.* 2008), provided that appropriate comparative databases are available.

Accepted Article

Potential for evolutionary studies

In the future, sequence alignments resulting from targeted mapping against reference nuclear and mtDNA sequences could be used to reconstruct the evolutionary and demographic trajectories of molluscs (Leonardi *et al.* 2016). Sampling shells in time-series would make it possible to investigate the impact of past environmental changes (e.g., temperature, salinity, pollutants, pathogenic outbreaks) that can be tracked from the same specimens through a range of palaeoenvironmental proxies (Chauvaud *et al.* 2005; Thébault *et al.* 2009; Lartaud *et al.* 2010; Szymanek 2011; Penkman *et al.* 2011; Butler *et al.* 2013; Pérez-Mayol *et al.* 2014; Reynolds *et al.* 2016). This information could in turn help predict mollusc demographic and adaptive responses to current and future environmental changes, such as ocean acidification, which, caused by the dissolution of carbon dioxide excess in the atmosphere, can hinder shell formation in molluscs (Parker *et al.* 2013; Gazeau *et al.* 2013). Additionally, ancient shell DNA could be applied to the conservation of endangered mollusc species. As previously done with museum material from mammalian endangered species (Der Sarkissian *et al.* 2015), temporally sampled DNA data could help track changes in genetic diversity through time, assess population genetic viability and identify specific candidate adaptive alleles in order to inform and improve conservation management. Another interesting application of marine mollusc shell DNA time-series is the reconstruction of past human diets from shell midden assemblages at archaeological sites, thereby shedding light on resource usage in ancient human populations and their adaptation to changing environments (Sanger 1993).

Potential for the studies of environmental and pathogenic bacteria

Our results, in particular the recovery of DNA likely belonging to *V. tapetis*, the etiological agent of BRD, also indicate the potential of ancient shell DNA to reconstruct the history of pathogenic diseases. BRD first appeared in Brittany in 1987 and has since caused high mortality in European aquacultures of Manila clams (Paillard *et al.* 1989, 1994; Borrego *et al.* 1996; Allam *et*

Accepted Article

al. 2000; Paillard *et al.* 2008). Interestingly, we found that DNA extracted from shells exhibiting small brown conchiolin deposits characteristic of BRD and collected in Brittany four years before the first BRD outbreak showed highest genetic affinity for the *V. tapetis* strain HH6087, while affinity for the *V. tapetis* strain RP2-3 was observed in shells post-dating the emergence of BRD in Brittany. These results tend to suggest that strains showing lower virulence than strains responsible for the BRD epizooty were present in Brittany before 1987. Further work, including genome sequencing of additional strains, characterization of virulence genes, and DNA analyses of the whole BRD time series, are however required to refine the history of BRD. This will more generally also help identify the genomic changes underlying the mechanisms of co-evolution between *V. tapetis* and Manila clams, which have developed recovery processes through the deposition of a mineral layer covering brown deposits and, hence, a resistance capacity to BRD (Trinkler *et al.* 2010a; b, 2011a; b; Jeffroy *et al.* 2013). Our results also suggest that DNA analyses of shell material hold the potential for new tools to detect and monitor diseases in aquaculture.

Finally, we detected DNA from various *Vibrio* species in shells. *Vibrio* is a diverse genus of the marine environment (Colwell 1996), which has become increasingly abundant in the North Atlantic within the last 50 years, following the rise of sea surface temperatures (Vezzulli *et al.* 2016; Baker-Austin *et al.* 2016). This spread has also been associated with an increase in the prevalence of severe human diseases caused by *V. cholerae*, *V. parahaemolyticus* and *V. vulnificus* (Vezzulli *et al.* 2016; Baker-Austin *et al.* 2016). Screening for *Vibrio* species in shell DNA could help detect similar trends in wider temporal and geographical ranges, thus providing invaluable information to predict future human disease (re-) emergence as a consequence of global warming.

Accepted Article

Conclusion

By recovering DNA from ancient marine mollusc shells as old as ~7,000 years, we validated mollusc shells as a new ancient DNA reservoir. This opens to (meta)genomic analyses the important palaeontological and archaeological record of marine mollusc shells, found in e.g., sediment cores, shell lags on the seabed, shell middens and museum collections. Future studies thus hold the potential to provide invaluable information to research in archaeology, evolutionary biology, marine community ecology and climate change. DNA analyses of mollusc shells could also be applied to study living communities, including the development of conservation programmes for threatened species, the genetic monitoring of economically important mollusc species, as well as the biological surveillance of aquaculture and early detection of pathogen outbreaks.

ACKNOWLEDGMENTS

We thank Tom Schiøtte and Martin Vinther Sørensen at the Zoological Museum of Copenhagen for providing historical shell samples from the Invertebrate Collection. We thank Adeline Bidault for *Vibrio* DNA extraction, Kristian Hanghøj for technical assistance, Mikkel Schubert and Gabriel Renaud for fruitful discussions, the PALEOMIX group and the staff of the Danish National High-Throughput DNA Sequencing Centre for support. This work was supported by the Danish Council for Independent Research, Natural Sciences (FNU, 4002-00152B); the Danish National Research Foundation (DNRF94); the EPT PROXACHEOBIO from Université Européenne de Bretagne (2010-2012); the APEGE initiative PaleoCOO of the Centre National de la Recherche Scientifique; the cluster of excellence LabexMER (ANR-10-LABX-19; METHOMOL) under the program “Investissements d’Avenir”; the UK Natural Environment Research Council (NE/H023356/1); the EU Marie Curie ARAMACC Initial Training Network (FP7-PEOPLE-2013-ITN 604802); the “Chaires d'Attractivité 2014” IDEX, University of Toulouse, France.

Accepted Article

REFERENCES

Allam B, Paillard C, Howard A, Le Pennec M (2000) Isolation of the pathogen *Vibrio tapetis* and defense parameters in brown ring diseased Manila clams *Ruditapes philippinarum* cultivated in England. *Diseases of Aquatic Organisms*, **41**, 105–113.

Allam B, Pales Espinosa E, Tanguy A *et al.* (2014) Transcriptional changes in Manila clam (*Ruditapes philippinarum*) in response to Brown Ring Disease. *Fish & Shellfish Immunology*, **41**, 2–11.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Andersen SH (2008) A report on recent excavations at the shell midden of Havnø in Denmark. *Mesolithic Miscellany*, **19**, 3–6.

Artigaud S, Richard J, Thorne MAS *et al.* (2015) Deciphering the molecular adaptation of the king scallop (*Pecten maximus*) to heat stress using transcriptomics and proteomics. *BMC genomics*, **16**, 988.

Artigaud S, Thorne MAS, Richard J *et al.* (2014) Deep sequencing of the mantle transcriptome of the great scallop *Pecten maximus*. *Marine Genomics*, **15**, 3–4.

Ávila-Arcos MC, Sandoval-Velasco M, Schroeder H *et al.* (2015) Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods in Ecology and Evolution*, **6**, 725–734.

Baker-Austin C, Trinanes J, Gonzalez-Escalona N, Martinez-Urtaza J (2016) Non-Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends in Microbiology*.

Bankevich A, Nurk S, Antipov D *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, **19**, 455–477.

Beniash E (2011) Biominerals—hierarchical nanocomposites: the example of bone. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, **3**, 47–69.

Bernstein BE, Meissner A, Lander ES (2007) The Mammalian Epigenome. *Cell*, **128**, 669–681.

Bidault A, Richard GG, Le Bris C, Paillard C (2015) Development of a Taqman real-time PCR assay for rapid detection and quantification of *Vibrio tapetis* in extrapallial fluids of clams. *PeerJ*, **3**, e1484.

Borrego JJ, Castro D, Luque A *et al.* (1996) *Vibrio tapetis* sp. nov., the Causative Agent of the Brown Ring Disease Affecting Cultured Clams. *International Journal of Systematic and Evolutionary Microbiology*, **46**, 480–484.

Bos KI, Schuenemann VJ, Golding GB *et al.* (2011) A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature*, **478**, 506–510.

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**, 59–60.

Butler PG, Wanamaker Jr. AD, Scourse JD, Richardson CA, Reynolds DJ (2013) Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on

Accepted Article

growth increments in the bivalve *Arctica islandica*. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **373**, 141–151.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC bioinformatics*, **10**, 421.

Champagne A, Aoustin D, Dupont C (2012) La citadelle de Brouage et la dynamique paléoenvironnementale du marais charentais : l'apport de la malacologie et de la palynologie. In: *Bilan Scientifique Régional 2011 de Poitou-Charentes.*, pp. 294–303. Service Régional de l'Archéologie, Poitiers.

Chauvaud L, Lorrain A, Dunbar RB *et al.* (2005) Shell of the Great Scallop *Pecten maximus* as a high-frequency archive of paleoenvironmental changes. *Geochemistry, Geophysics, Geosystems*, **6**, Q08001.

Chauvaud L, Patry Y, Jolivet A *et al.* (2012) Variation in Size and Growth of the Great Scallop *Pecten maximus* along a Latitudinal Gradient. *PLOS ONE*, **7**, e37717.

Chowdhury PR, Boucher Y, Hassan KA *et al.* (2011) Genome Sequence of *Vibrio rotiferianus* Strain DAT722□. *Journal of Bacteriology*, **193**, 3381–3382.

Colwell RR (1996) Global Climate and Infectious Disease: The Cholera Paradigm. *Science*, **274**, 2025–2031.

Crosetto N, Bienko M, van Oudenaarden A (2015) Spatially resolved transcriptomics and beyond. *Nature Reviews Genetics*, **16**, 57–66.

Cruz-Dávalos DI, Llamas B, Gaunitz C *et al.* (2016) Experimental conditions improving in solution target enrichment for ancient DNA. *Molecular Ecology Resources*.

Dabney J, Knapp M, Glocke I *et al.* (2013a) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15758–15763.

Dabney J, Meyer M, Paabo S (2013b) Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology*, **5**, a012567–a012567.

Der Sarkissian C, Ermini L, Schubert M *et al.* (2015) Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Current Biology*, **25**, 2577–2583.

Desriac F, Le Chevalier P, Brillet B *et al.* (2014) Exploring the hologenome concept in marine bivalvia: haemolymph microbiota as a pertinent source of probiotics for aquaculture. *FEMS microbiology letters*, **350**, 107–116.

Ermini L, Der Sarkissian C, Willerslev E, Orlando L (2015) Major transitions in human evolution revisited: a tribute to ancient DNA. *Journal of Human Evolution*, **79**, 4–20.

Fierer N, Leff JW, Adams BJ *et al.* (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, **109**, 21390–21395.

Gamba C, Hanghøj K, Gaunitz C *et al.* (2016) Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*, **16**, 459–469.

Gamba C, Jones ER, Teasdale MD *et al.* (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, **5**.

Accepted Article

Gazeau F, Parker LM, Comeau S *et al.* (2013) Impacts of ocean acidification on marine shelled molluscs. *Marine Biology*, **160**, 2207–2245.

Geist J, Wunderlich H, Kuehn R (2008) Use of mollusc shells for DNA-based molecular analyses. *Journal of Molluscan Studies*, **74**, 337–343.

Ghiselli F, Milani L, Chang PL *et al.* (2012) De Novo assembly of the Manila clam *Ruditapes philippinarum* transcriptome provides new insights into expression bias, mitochondrial doubly uniparental inheritance and sex determination. *Molecular Biology and Evolution*, **29**, 771–786.

Gissinger B (2011) La Gripperie-Saint-Symphorien, Le Cimetière 2. In: *Bilan Scientifique Régional*, pp. 92–93. Direction Régionale des Affaires Culturelles (DRAC) Poitou-Charentes, Service Régional d' Archéologie (SRA), Ministère de la Culture et de la Communication, Poitiers.

Gokhman D, Lavi E, Prüfer K *et al.* (2014) Reconstructing the DNA Methylation Maps of the Neandertal and the Denisovan. *Science*, **344**, 523–527.

Gröcke DR, Gillikin DP (2008) Advances in mollusc sclerochronology and sclerochemistry: tools for understanding climate and environment. *Geo-Marine Letters*, **28**, 265–268.

Hanghøj K, Seguin A, Schubert M *et al.* (2016) Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Molecular Biology and Evolution*, msw184.

Harney E, Dubief B, Boudry P *et al.* (2016) De novo assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome. *Marine Genomics*, **28**, 11–16.

Hiebenthal C, Philipp EER, Eisenhauer A, Wahl M (2012) Interactive effects of temperature and salinity on shell formation and general condition in Baltic Sea *Mytilus edulis* and *Arctica islandica*. *Aquatic Biology*, **14**, 289–298.

Huang W, Li L, Myers JR, Marth GT (2012) ART: a next-generation sequencing read simulator. *Bioinformatics*, **28**, 593–594.

Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Research*, **17**, 377–386.

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, **21**, 1552–1560.

Jeffroy F, Brulle F, Paillard C (2013) Differential expression of genes involved in immunity and biomineralization during Brown Ring Disease development and shell repair in the Manila clam, *Ruditapes philippinarum*. *Journal of Invertebrate Pathology*, **113**, 129–136.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, **29**, 1682–1684.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Accepted Article

Large J-M, Birocheau P (2004) *Les Châtelliers du Vieil-Auzay (Vendée). Une archéologie d'un site exceptionnel de la Préhistoire récente*. Groupe Vendéen d'Etudes Préhistoriques, La Roche-sur-Yon.

Lartaud F, Emmanuel L, de Rafaelis M, Pouvreau S, Renard M (2010) Influence of food supply on the $\delta^{13}C$ signature of mollusc shells: implications for palaeoenvironmental reconstitutions. *Geo-Marine Letters*, **30**, 23–34.

Le Goff L, Dupont C (2015) Consommation de coquillages du Moyen Âge au début de l'époque moderne sur le littoral charentais : les exemples de Fontdouce et de La Gripperie-Saint-Symphorien. *Aquitania*, **31**, 373–400.

Leonardi M, Librado P, Der Sarkissian C *et al.* (2016) Evolutionary Patterns and Processes: Lessons from Ancient DNA. *Systematic Biology*, syw059.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Lokmer A, Goedknegt MA, Thieltges DW *et al.* (2016) Spatial and Temporal Dynamics of Pacific Oyster Hemolymph Microbiota across Multiple Scales. *Microbial Symbioses*, 1367.

Lokmer A, Wegner KM (2015) Hemolymph microbiome of Pacific oysters in response to temperature, temperature stress and infection. *The ISME Journal*, **9**, 670–682.

Louvel G, Der Sarkissian C, Hanghøj K, Orlando L (2016) metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data. *Molecular Ecology Resources*, n/a-n/a.

MacHugh DE, Larson G, Orlando L (2016) Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual Review of Animal Biosciences*.

Marin F, Le Roy N, Marie B (2012) The formation and mineralization of mollusk shell. *Frontiers in Bioscience (Scholar Edition)*, **4**, 1099–1125.

Meisterhans G, Raymond N, Girault E *et al.* (2015) Structure of Manila Clam (*Ruditapes philippinarum*) Microbiota at the Organ Scale in Contrasting Sets of Individuals. *Microbial Ecology*, **71**, 194–206.

Metzker ML (2010) Sequencing technologies — the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Meyer M, Arsuaga J-L, de Filippo C *et al.* (2016) Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, **531**, 504–507.

Meyer M, Fu Q, Aximu-Petri A *et al.* (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, **505**, 403–406.

Milan M, Coppe A, Reinhardt R *et al.* (2011) Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum*: genomic tools for environmental monitoring. *BMC Genomics*, **12**, 234.

Moreira R, Balseiro P, Planas JV *et al.* (2012a) Transcriptomics of in vitro immune-stimulated hemocytes from the Manila clam *Ruditapes philippinarum* using high-throughput sequencing. *PloS One*, **7**, e35009.

Moreira R, Balseiro P, Romero A *et al.* (2012b) Gene expression analysis of clams *Ruditapes philippinarum* and *Ruditapes decussatus* following bacterial infection yields molecular

Accepted Article

insights into pathogen resistance and immunity. *Developmental and Comparative Immunology*, **36**, 140–149.

Moreira R, Pereiro P, Canchaya C *et al.* (2015) RNA-Seq in *Mytilus galloprovincialis*: comparative transcriptomics and expression profiles among different tissues. *BMC Genomics*, **16**, 728.

Murgarella M, Puiu D, Novoa B *et al.* (2016) A First Insight into the Genome of the Filter-Feeder Mussel *Mytilus galloprovincialis*. *PLOS ONE*, **11**, e0151561.

Mustafa GA, Abd-Elgawad A, Ouf A, Siam R (2016) The Egyptian Red Sea coastal microbiome: A study revealing differential microbial responses to diverse anthropogenic pollutants. *Environmental Pollution (Barking, Essex: 1987)*, **214**, 892–902.

Nys Y, Gautron J, Garcia-Ruiz JM, Hincke MT (2004) Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. *Comptes Rendus Palevol*, **3**, 549–562.

Orlando L, Gilbert MTP, Willerslev E (2015) Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics*, **16**, 395–408.

Orlando L, Ginolhac A, Zhang G *et al.* (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, **499**, 74–78.

Orlando L, Willerslev E (2014) An epigenetic window into the past? *Science*, **345**, 511–512.

Oskam CL, Haile J, McLay E *et al.* (2010) Fossil avian eggshell preserves ancient DNA. *Proceedings. Biological Sciences / The Royal Society*, **277**, 1991–2000.

Paillard C, Ashton-Alcox KA, Ford SE (1996) Changes in hemolymph and extrapallial fluid parameters in the american oyster, *Crassostrea virginica* affected by the Juvenile oyster disease. *Aquatic Living Resources*, **9**, 145–148.

Paillard C, Korsnes K, Le Chevalier P *et al.* (2008) Vibrio tapetis-like strain isolated from introduced Manila clams *Ruditapes philippinarum* showing symptoms of brown ring disease in Norway. *Diseases of Aquatic Organisms*, **81**, 153–161.

Paillard C, Maes P (1995) The Brown Ring Disease in the Manila Clam, *Ruditapes philippinarum*. *Journal of Invertebrate Pathology*, **65**, 101–110.

Paillard C, Maes P, Oubella R (1994) Brown ring disease in clams. *Annual Review of Fish Diseases*, **4**, 219–240.

Paillard C, Percelay L, Le Pennec M, Le Picard D (1989) Origine pathogène de l'"anneau brun" chez *Tapes philippinarum* (Mollusque, bivalve). *Comptes rendus de l'Académie des sciences*, **309**, 235–241.

Paillard C, Roux FL, Borrego JJ (2004) Bacterial disease in marine bivalves, a review of recent studies: Trends and evolution. *Aquatic Living Resources*, **17**, 477–498.

Parker LM, Ross PM, O'Connor WA *et al.* (2013) Predicting the Response of Molluscs to the Impact of Ocean Acidification. *Biology*, **2**, 651–692.

Pauletto M, Milan M, Moreira R *et al.* (2014) Deep transcriptome sequencing of *Pecten maximus* hemocytes: a genomic resource for bivalve immunology. *Fish & Shellfish Immunology*, **37**, 154–165.

Accepted Article

Pedersen MW, Overballe-Petersen S, Ermini L *et al.* (2015) Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **370**, 20130383.

Pedersen JS, Valen E, Velazquez AMV *et al.* (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Research*, **24**, 454–466.

Penkman KEH, Preece RC, Bridgland DR *et al.* (2011) A chronological framework for the British Quaternary based on *Bithynia* opercula. *Nature*, **476**, 446–449.

Pérez-Mayol S, Blasco J, Tornero V *et al.* (2014) Are the shells of *Scrobicularia plana* useful for monitoring trace metal pollution events? *Journal of Environmental Biology / Academy of Environmental Biology, India*, **35**, 9–17.

Ponder WF, Lindberg DR (2008) *Phylogeny and Evolution of the Mollusca*. University of California Press.

Reid HI, Duncan HL, Laidler LA, Hunter D, Birkbeck TH (2003) Isolation of *Vibrio tapetis* from cultivated Atlantic halibut (*Hippoglossus hippoglossus* L.). *Aquaculture*, **221**, 65–74.

Reimer PJ, Bard E, Bayliss A et al. (2013) IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, **55**, 1869–1887.

Reynolds DJ, Scourse J, Halloran P *et al.* (2016) Annually-resolved North Atlantic marine climate over the last millenium. *Nature Communications*. *In press*.

Sanger D (1993) Deciphering a shell midden. *Geoarchaeology*, **8**, 165–167.

Schroeder H, Ávila-Arcos MC, Malaspinas A-S *et al.* (2015) Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences*, **112**, 3669–3673.

Schubert M, Ermini L, Der Sarkissian C *et al.* (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, **9**, 1056–1082.

Schubert M, Ginolhac A, Lindgreen S *et al.* (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, **13**, 178.

Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, **9**, 88.

Segata N, Boernigen D, Tickle TL *et al.* (2013) Computational meta'omics for microbial community studies. *Molecular Systems Biology*, **9**, 666.

Segata N, Waldron L, Ballarini A *et al.* (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, **9**, 811–814.

Seguin-Orlando A, Schubert M, Clary J *et al.* (2013) Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PloS One*, **8**, e78575.

Slatkin M, Racimo F (2016) Ancient DNA and human history. *Proceedings of the National Academy of Sciences*, **113**, 6380–6387.

Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution*, **45**, 203–217.

Accepted Article

Sunagawa S, Coelho LP, Chaffron S *et al.* (2015) Structure and function of the global ocean microbiome. *Science*, **348**, 1261359.

Szymanek M (2011) Climate oscillations of the Holsteinian (Mazovian) Interglacial recorded in shell morphometry of *Viviparus diluvianus* (Kunth, 1865) from eastern Poland. *Quaternary International*, **241**, 143–159.

Thébault J, Chauvaud L, L'Helguen S *et al.* (2009) Barium and molybdenum records in bivalve shells: Geochemical proxies for phytoplankton dynamics in coastal environments? *Limnology and Oceanography*, **54**, 1002–1014.

Trinkler N, Bardeau J-F, Marin F *et al.* (2011a) Mineral phase in shell repair of Manila clam *Venerupis philippinarum* affected by brown ring disease. *Diseases of Aquatic Organisms*, **93**, 149–162.

Trinkler N, Guichard N, Labonne M *et al.* (2011b) Variability of shell repair in the Manila clam *Ruditapes philippinarum* affected by the Brown Ring Disease: a microstructural and biochemical study. *Journal of Invertebrate Pathology*, **106**, 407–417.

Trinkler N, Labonne M, Marin F *et al.* (2010a) Clam shell repair from the brown ring disease: a study of the organic matrix using confocal Raman micro-spectrometry and WDS microprobe. *Analytical and Bioanalytical Chemistry*, **396**, 555–567.

Trinkler N, Sinquin G, Querne J, Paillard C (2010b) Resistance to Brown Ring Disease in the Manila clam, *Ruditapes philippinarum*: a study of selected stocks showing a recovery process by shell repair. *Journal of Invertebrate Pathology*, **104**, 8–16.

Truong DT, Franzosa EA, Tickle TL *et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, **12**, 902–903.

Vezzulli L, Grande C, Reid PC *et al.* (2016) Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic. *Proceedings of the National Academy of Sciences*, **113**, E5062–E5071.

Vilstrup JT, Seguin-Orlando A, Stiller M *et al.* (2013) Mitochondrial phylogenomics of modern and ancient equids. *PloS One*, **8**, e55950.

Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR (1998) Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*, **105**, 539–543.

Zbawicka M, Wenne R, Burzyński A (2014) Mitogenomics of recombinant mitochondrial genomes of Baltic Sea *Mytilus* mussels. *Molecular Genetics and Genomics*, **289**, 1275–1287.

Zhang G, Fang X, Guo X *et al.* (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, **490**, 49–54.

Accepted Article

DATA ACCESSIBILITY

All sequencing data generated in this study are available from the European Nucleotide Archive (project PRJEB20113).

AUTHOR CONTRIBUTIONS:

CDe and LO conceived and coordinated the study. CDe performed molecular and computational analyses, with input from VP, MP and LO. VP, CDu, TS, PCI, CP and LO provided samples and reagents. CDe and LO wrote the article.

FIGURES

Figure 1. Preliminary tests: taxon detection from bleached and non-bleached modern shells.

(A) Metagenomic analyses: number of BLAST *nr* database hits, normalized, for each sample, by the total number of reads retained after quality filtering (retained reads). (B) Targeted mapping: coverage is calculated considering the total number of unique high-quality (Mapping Quality, MQ≥30) reads, normalized, for each sample, by the total number of retained reads. "I^-B^-", non-infected non-bleached fragment; "I^+B^-", infected non-bleached fragment; "I^-B^+", non-infected bleached fragment; "I^+B^+", infected bleached fragment.

Figure 2. Taxon detection by metagenomic analyses.

The colour gradient represents the proportion of the total number of assigned reads that were assigned to taxa for each taxonomic level.

Figure 3. Microbial diversity in shells, seawater, soil and human-associated samples.

PCoA of Bray-Curtis distances between genus-level microbial profiles. (A) Considering all taxa. (B) Considering only taxa absent from EBCs and after re-adjusting relative abundances.

Accepted Article

Figure 4. Mollusc taxon identification.

The colour gradient represents the coverage, scaled by sample, obtained when mapping against each (A) nuclear, (B) mtDNA genome, and (C) mtDNA barcode reference sequence, considering unique high-quality (MQ≥30) reads, and normalized, for each sample, by the sequencing effort. For each comparative database and sample, the yellow dot indicates the reference sequence for which the maximal depth-of-coverage was obtained, the value of which is specified at the top of the corresponding panels. "*" signifies depths-of-coverage <0.1.

Figure 5. *Vibrio* species identification.

The colour gradient represents the coverage obtained when mapping against each *Vibrio* species reference genome sequence, considering unique high-quality (MQ≥30) reads, and normalized, for each sample, by the total number of reads retained after adapter trimming/quality filtering. For each BRD sample, the black dot indicates the reference sequence for which the maximal depth-of-coverage was obtained. For species other than *Vibrio tapetis*, results for one representative genome per species are shown (see Figure S7 for mapping results against the full *Vibrio* panel).

Figure 6. Post-mortem DNA damage patterns.

For hav_Mussel, reads mapping to the *Mytilus galloprovincialis* nuclear transcript reference sequences were considered to describe (A) fragment size distribution, (B) nucleotide misincorporation, and (C) DNA fragmentation shown within the first 10 bp from read ends (post- quality trimming). For LAN1p_Clam and ROS-B_Scallop, the reference sequences to which HTS reads were mapped are indicated above each plot. Nucleotide misincorporations are shown along the first ten read positions in (D) and (E), fragment size distributions in (F)-(J). Only unique high-quality (MQ≥30) reads were considered and mapDamage v.2 was run on the full alignment (100,0000 iterations).

Accepted Article

Table 1. Sample information and sequencing statistics.

Sample name	Species name	Geographical location	Date/Age	Inner shell microstructure	Shell diagnostic	#Sequences	#Retained
MODERN							
POS_Clam	*Ruditapes philippinarum*	Landéda, France	1988 AD	HA	Recovered after acute BRD	36,371,525	33,530,237
KOR_Clam		Cheju, Korea	2003 AD	HA	Acute BRD	35,083,934	34,501,909
AKKp_Clam		Akkeshi, Japan	2012 AD	HA	Abnormal brown deposit	9,709,393	9,640,133
MATp_Clam		Matsusaka, Japan	2012 AD	HA	Weak BRD	18,091,072	18,001,838
SELp_Clam		Seloy, Norway	2003 AD	HA	Weak BRD	8,455,559	8,373,378
LAN1p_Clam		Landéda, France	1983 AD	HA	Brown deposit	10,499,178	10,402,596
LAN2p_Clam		Landéda, France	1983 AD	HA	Brown deposit	13,812,143	13,667,939
LAN3p_Clam		Landéda, France	1983 AD	HA	Brown deposit	7,981,394	7,925,697
MURp_Clam		Murano, Italy	1990 AD	HA	Weak BRD	9,932,924	9,847,558
NEG_Clam		Landéda, France	1988 AD	HA	Healthy	50,373,801	50,694,338
BRE1_Abalone	*Haliotis tuberculata*	Brest, France	2003 AD	NA	Healthy	7,506,486	7,447,143
BRE2_Abalone		Brest, France	2005 AD	NA	Healthy	10,795,266	10,739,804
BRE3m_Abalone		Brest, France	2003 AD	NA	Healthy	9,749,078	9,675,009
BRE3p_Abalone		Brest, France	2003 AD	NA	Acute brown deposit and boring	9,184,721	9,123,253
LEF-B_Scallop	*Pecten maximus*	Le Faou, France	1960-1970 AD	FC	Healthy	20,565,155	20,104,424
ROS-B_Scallop		Roscanvel, France	1997 AD	FC	Healthy	7,080,764	7,040,427

Accepted Article

BRE5m_Oyster	*Crassostrea sp.*	Brest, France	2005-2007 AD	FC	Healthy	11,484,012	11,430,195
BRE5p_Oyster		Brest, France	2005-2007 AD	FC	Acute brown deposit and boring	7,724,821	7,686,285
ANCIENT							
hav_Mussel	*Mytilus sp.*	Havnø, Denmark	Late Mesolithic ~7,000 yBP[1]	NA	Healthy	11,183,672	10,515,059
auzay1B_Clam	*Ruditapes decussatus*	Les Châtelliers du Vieil-Auzay, France	Middle Neolithic 6th mill. BP[2]	NA	Healthy	15,240,748	15,030,940
auzay2B_Clam		Les Châtelliers du Vieil-Auzay, France	Middle Neolithic 6th mill. BP[2]	NA	Healthy	11,095,140	10,913,477
auzay3B_Clam		Les Châtelliers du Vieil-Auzay, France	Middle Neolithic 6th mill. BP[2]	NA	Healthy	13,565,581	13,414,246
auzay4B_Clam		Les Châtelliers du Vieil-Auzay, France	Middle Neolithic 6th mill. BP[2]	NA	Healthy	12,191,311	12,008,593
gripB_Clam		La Gripperie-Saint-Symphorien, France	7th cent. AD[3,4]	NA	Healthy	10,555,462	10,484,321
lmc1B_Clam		La Maison Champlain, Brouage, France	16-17th cent. AD[5]	NA	Healthy	12,983,981	12,888,532
lmc2B_Clam		La Maison Champlain, Brouage, France	16-17th cent. AD[5]	NA	Healthy	10,911,865	10,811,007
lmc3B_Clam		La Maison Champlain, Brouage, France	16-17th cent. AD[5]	NA	Healthy	17,964,100	17,845,554
nis_Clam	*Venerupis corrugata*	Nissum Bredning, Denmark	1886 AD	NA	Perforations	8,386,634	8,301,630
thi_Clam		Thisted Bredning, Denmark	1909 AD	NA	Healthy	13,947,789	13,868,126
med_Abalone	*Haliotis tuberculata*	Mediterranean Sea	Before 1804 AD	NA	Healthy	9,823,899	9,699,275
gue_Abalone		Guernesey Island, UK	1832 AD	NA	Healthy	481,087	475,317
hel_Quahog	*Arctica islandica*	Hellebæk, Denmark	1886 AD	HA	Healthy	13,089,376	12,674,074
quahog2_Quahog		sine loc.	Before 1804 AD	HA	Healthy	18,871,776	18,494,781
ves_Quahog		Vestmanna, Faroe Islands	1901 AD	HA	Healthy	8,264,494	8,153,984
Extraction blank controls for samples:							
EBC7_Control	auzay1-4B_Clam, gripB_Clam, lmc1-3B_Clam					8,914,833	8,653,416

Accepted Article

EBC8_Control	hav_Mussel, gue_Abalone, med_Abalone, ves_Quahog	7,557,184	3,570,529
EBC9_Control	quahog2_Quahog, hel_Quahog	9,974,348	3,146,625
EBC11_Control	nis_Clam, thi_Clam	9,597,406	8,798,083
EBC12_Control	MATp_Clam and MURp_Clam	106,077	74,324

The suffix "p" alone in the sample name indicates that the shell fragment showed signs of infection; "m" no signs of infection.

"AD", Anno Domini; "yBP", years Before Present; "mill.", millenium; "cent.", century; “HA”, homogeneous aragonite; “NA”, nacreous aragonite; “FC”, foliated calcite; "BRD", Brown Ring Disease.

[1](Andersen 2008); [2](Large & Birocheau 2004); [3](Gissinger 2011); [4](Le Goff & Dupont 2015); [5](Champagne *et al.* 2012)

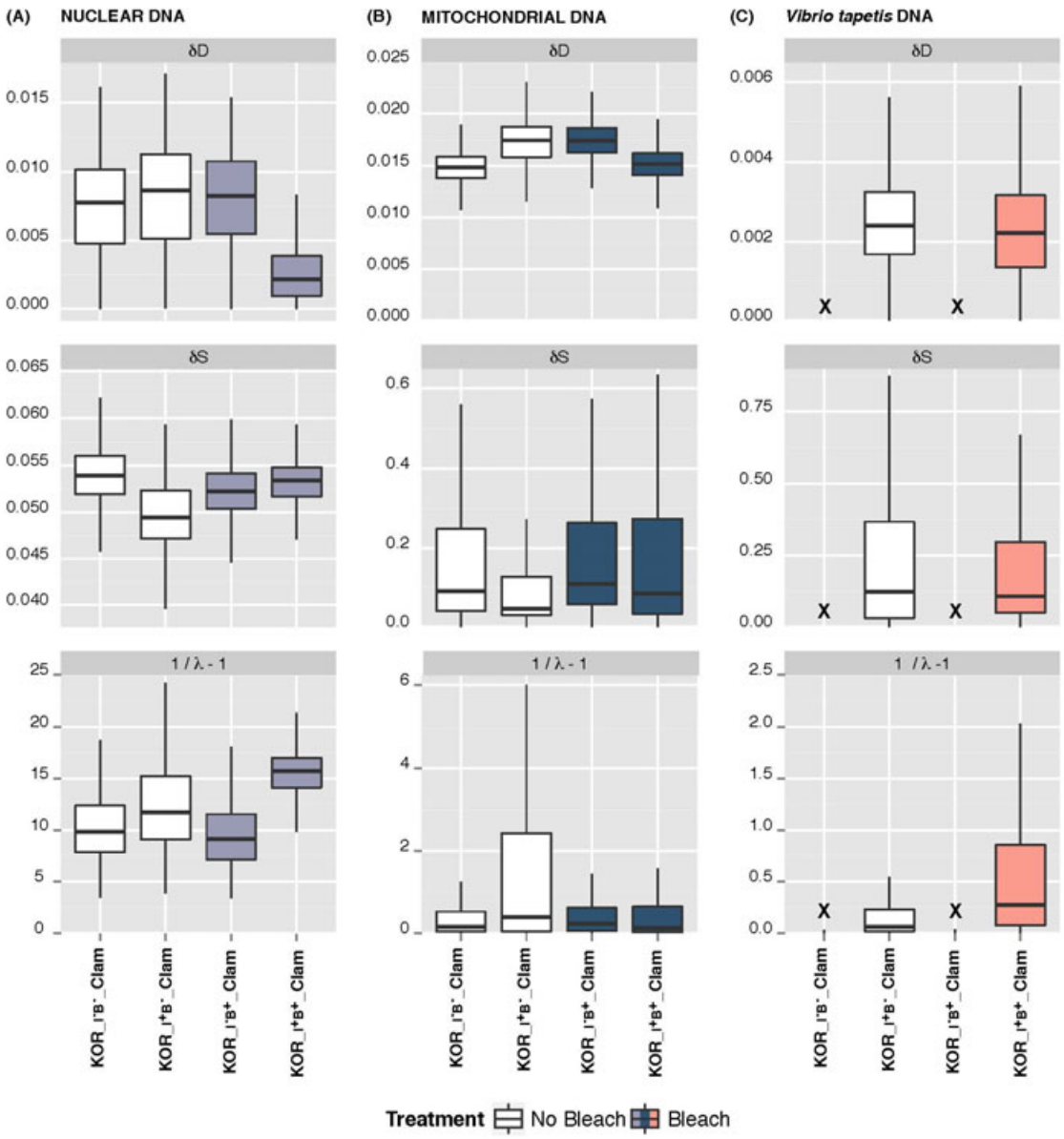
Accepted Article

Table 2. Mollusc taxon identification success rates.

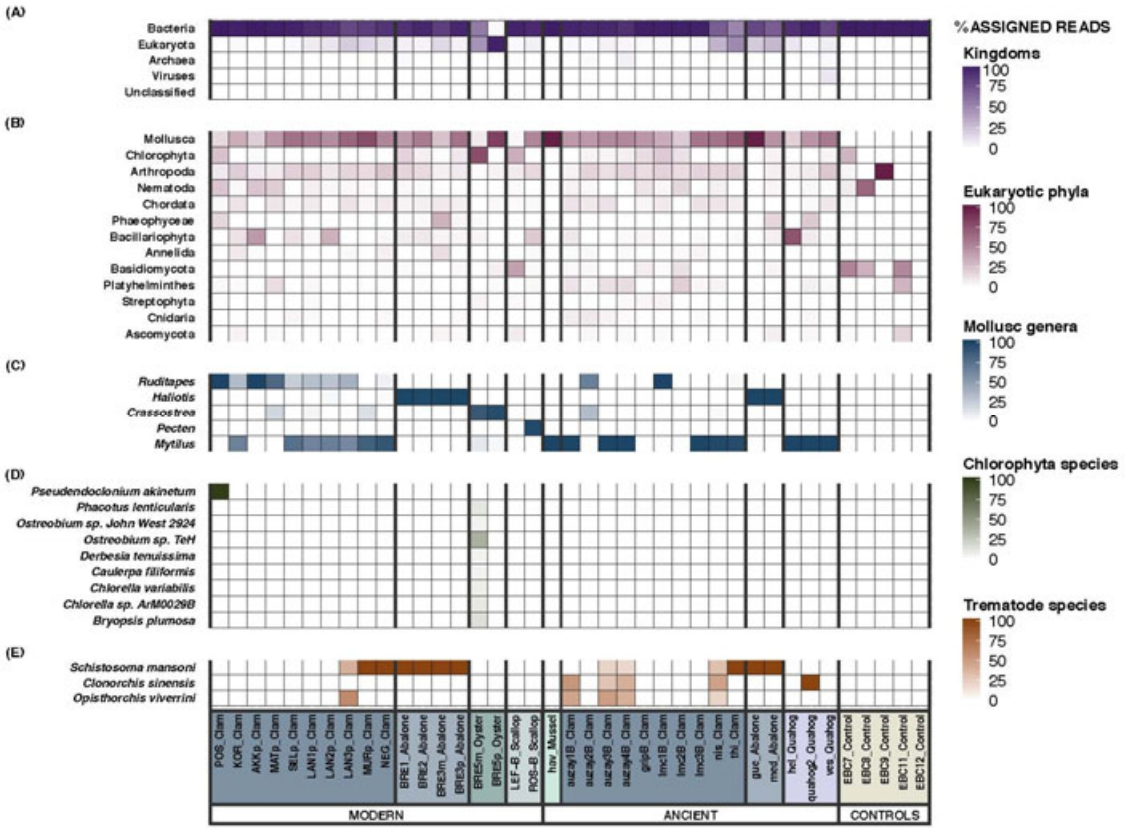
		Success rates across all samples % (sequenced only)			
Mollusc taxa	**Groups**	**Metagenomics**	**Nuclear DNA**	**mtDNA genomes**	**mtDNA barcodes**
Clams	**Modern**	33 (33)	100 (100)	90 (90)	100 (100)
Abalones		100 (100)	100 (100)	100 (100)	100 (100)
Oysters		33 (100)	33 (100)	33 (100)	33 (100)
Scallops		20 (50)	20 (50)	20 (50)	20 (50)
Mussels	**Ancient**	100 (100)	100 (100)	100 (100)	0 (0)
Clams		100 (100)	80 (80)	0 (0)	100 (100)
Abalones		20 (10)	100 (100)	100 (100)	100 (100)
Quahogs		100 (100)	0 (0)	100 (100)	100 (100)

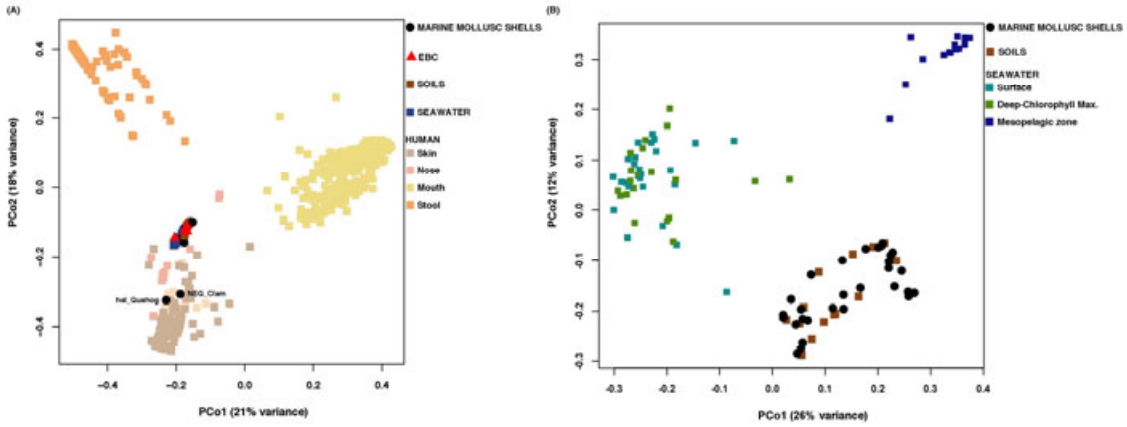
"mtDNA", mitochondrial DNA

Accepted Article

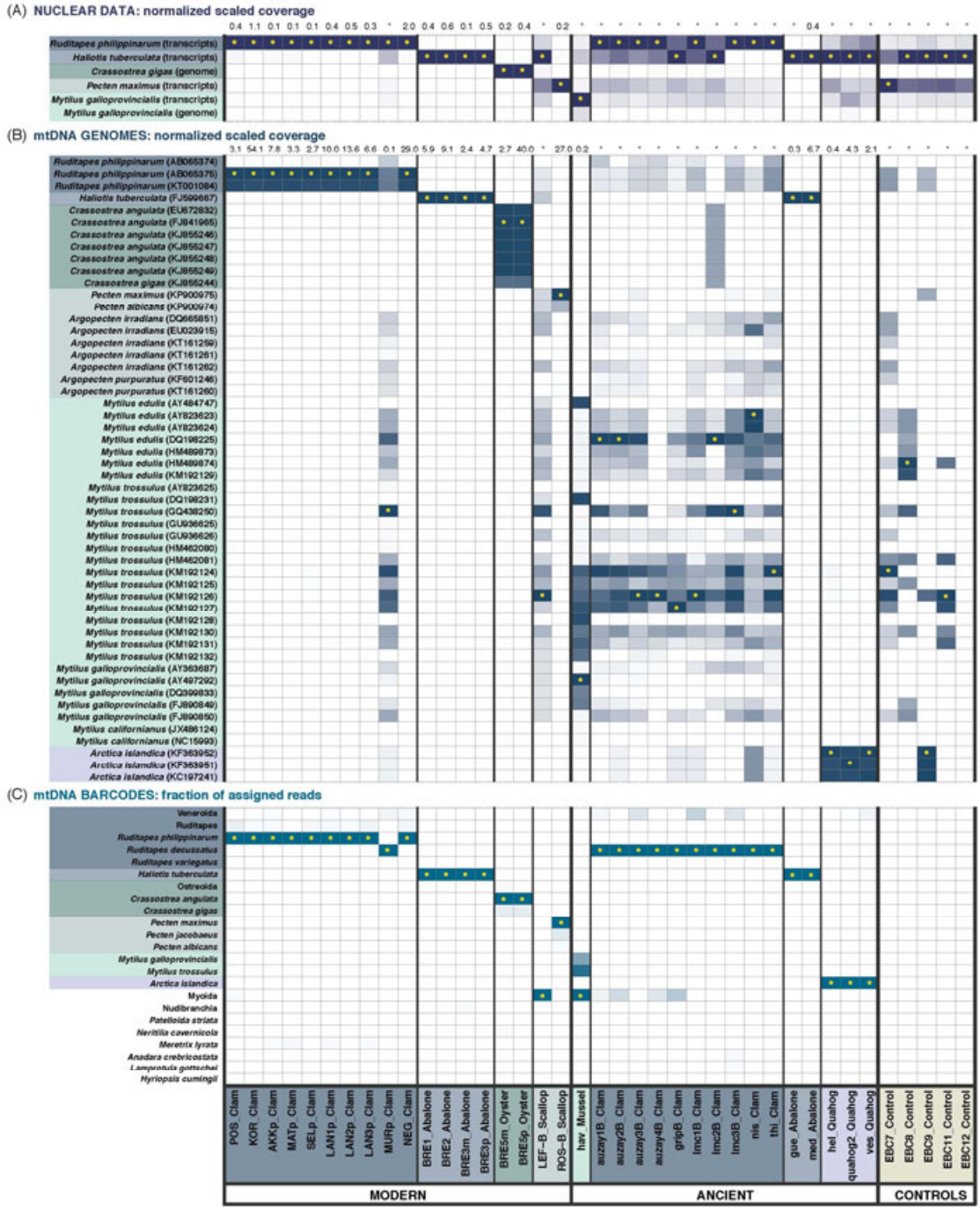


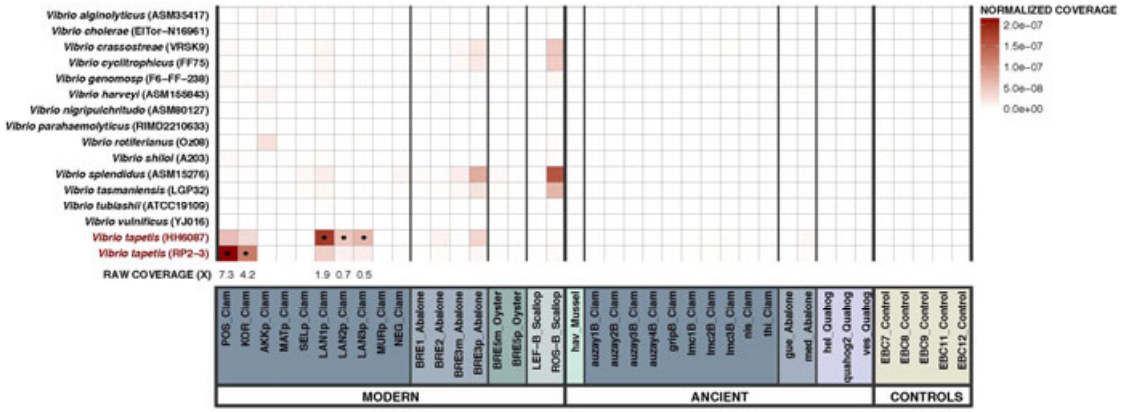
Accepted Article





Accepted Article





Accepted Article

Accepted Article

